Netflix Siblings

Taylor Blair

1/26/2021

#A brief intro

My sister and I have a long history of watching Netflix.

There are some minor outliers in the datasets, namely when we use each others accounts.

#Loading data and libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(openair)
library(readr)
```

```
my_history <- read_csv("~/Coding/R/sis-and-me-netflix/MyNetflixHistory.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Title = col_character(),
##   Date = col_character()
## )
```

```
sibling_history <- read_csv("~/Coding/R/sis-and-me-netflix/SisNetflixHistory.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Title = col_character(),
##   Date = col_character()
## )
```

#Data prep

##Date fixing

```
my_history$Date <- as.Date(my_history$Date,"%m/%d/%y")
sibling_history$Date <- as.Date(sibling_history$Date,"%m/%d/%y")
```

##Movie or TV series

```
my_history$type <- grepl(":", my_history$Title)
my_history$type <- my_history$type %>%
  replace(my_history$type==TRUE, "TV Series") %>%
  replace(my_history$type==FALSE, "Movie")

sibling_history$type <- grepl(":", sibling_history$Title)
sibling_history$type <- sibling_history$type %>%
  replace(sibling_history$type==TRUE, "TV Series") %>%
  replace(sibling_history$type==FALSE, "Movie")
```

##Series, Season and episode

```
my_history <- my_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")

sibling_history <- sibling_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
```

#Day of Week

```
day_of_week <- function(x){
  return (weekdays(x))
}

days <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")

my_history$day <- sapply(my_history$Date, day_of_week)
my_history$day <- factor(my_history$day,
                         levels = days)

sibling_history$day <- sapply(sibling_history$Date, day_of_week)
sibling_history$day <- factor(sibling_history$day,
                              levels = days)
```

#Compare dataset

```
my_history$individual <- "Tay"
sibling_history$individual <- "Sis"

merged <- rbind(my_history, sibling_history)
```

#Numerical analysis

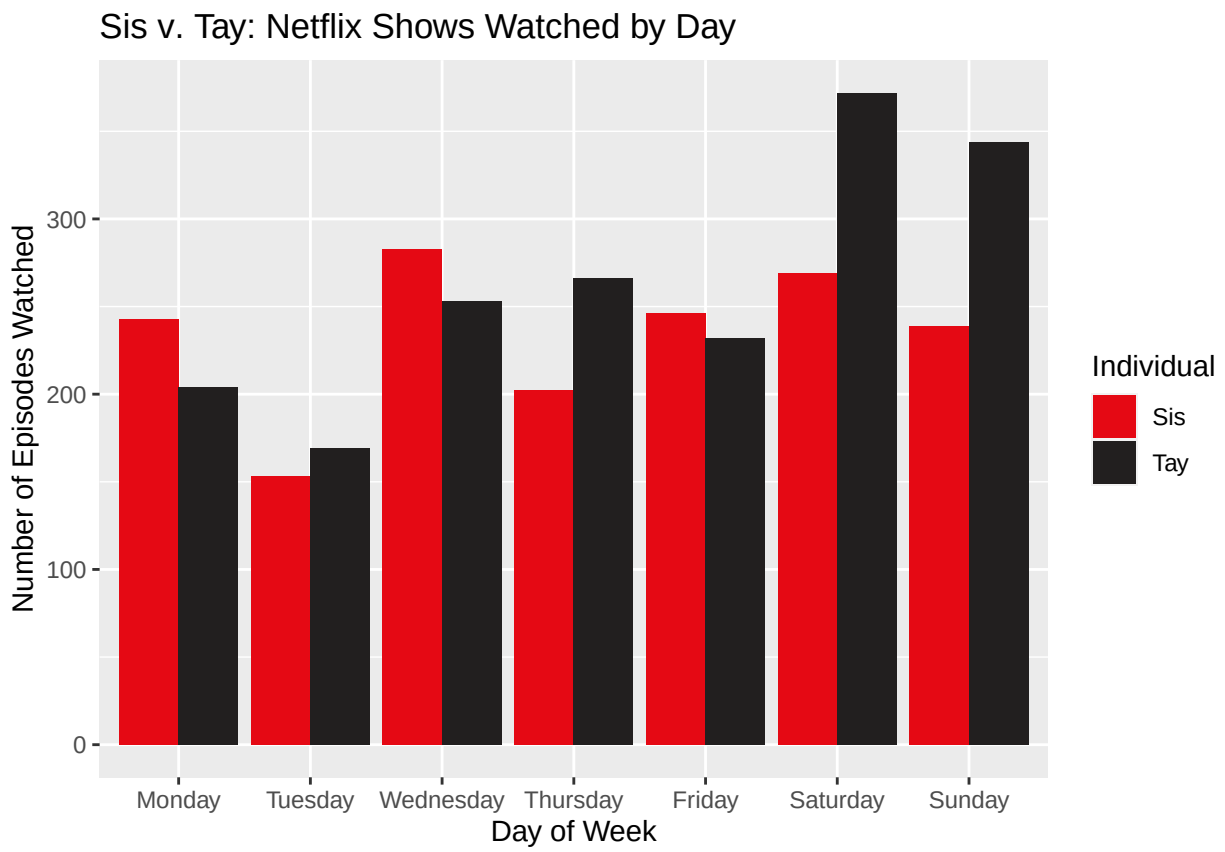
##Binge Watch pisodes

#Graphical Analysis

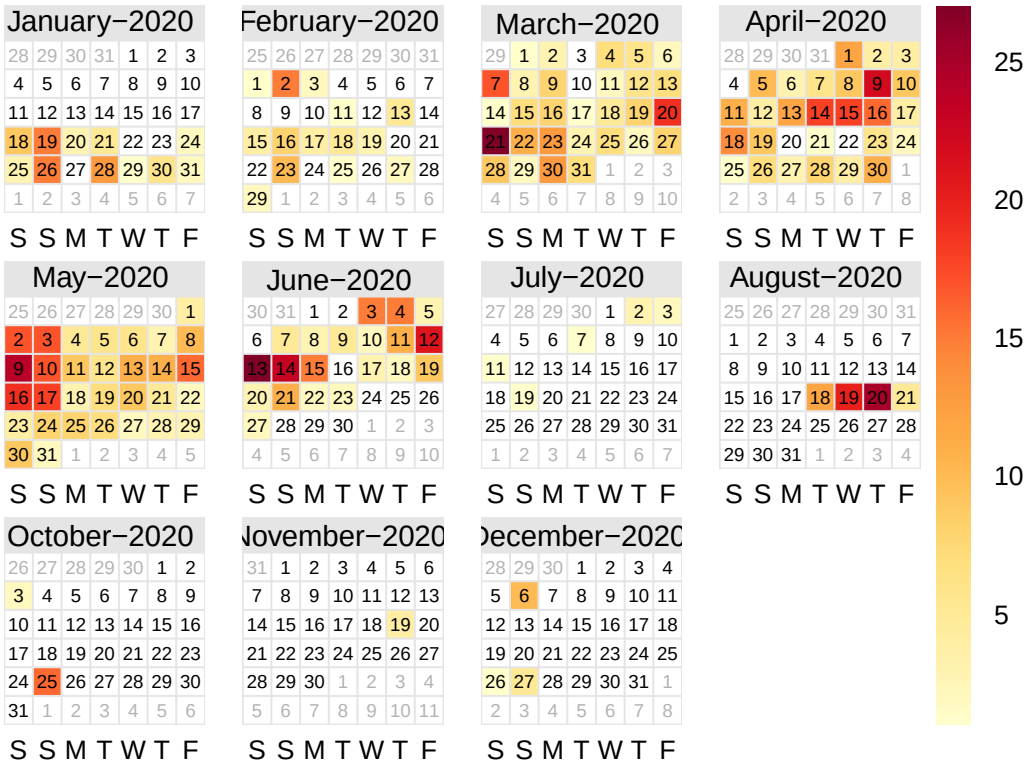
Looking at

```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs(title = "Sis v. Tay: Netflix Shows Watched by Day") +
  scale_fill_manual("Individual", values=c("#E50914","#221F1F")) +
```

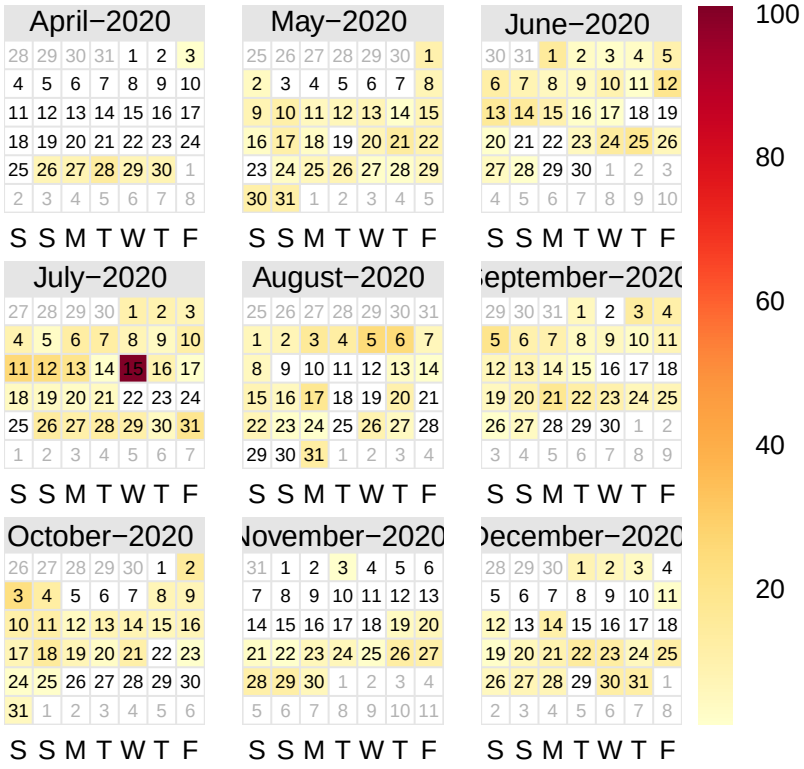
```
  xlab("Day of Week") +
  ylab("Number of Episodes Watched")
```




```
my_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020)
```




```
sibling_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020,
               month = c(1:12))
```


April−2020
28 29 30 31 1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 1
2 3 4 5 6 7 8
S S M T W T F
May−2020
25 26 27 28 29 30 1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30 31 1 2 3 4 5
S S M T W T F
June−2020
30 31 1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18 19
20 21 22 23 24 25 26
27 28 29 30 1 2 3
4 5 6 7 8 9 10
S S M T W T F
July−2020
27 28 29 30 1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31
1 2 3 4 5 6 7
S S M T W T F
August−2020
25 26 27 28 29 30 31
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 1 2 3 4
S S M T W T F
eptember−2020
29 30 31 1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 1 2
3 4 5 6 7 8 9
S S M T W T F
October−2020
26 27 28 29 30 1 2
3 4 5 6 7 8 9
10 11 12 13 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
31 1 2 3 4 5 6
S S M T W T F
November−2020
31 1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 1 2 3 4
5 6 7 8 9 10 11
S S M T W T F
December−2020
28 29 30 1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 31 1
2 3 4 5 6 7 8
S S M T W T F
100
80
60
40
20